Quality Estimation based Data Fusion in Wireless Sensor Networks

Frederik Hermans, Norman Dziengel, Jochen Schiller
Institute of Computer Science
Freie Universität Berlin
Berlin, Germany
Email: {hermans,dziengel,schiller}@inf.fu-berlin.de

Abstract

The main purpose of wireless sensor networks (WSNs) is to obtain information about their environment. However, WSNs often produce imprecise and incorrect sensor data, e.g. because of sensor failure or unreliable radio communication. We propose a system for WSN applications that allows to assess the quality of sensor data and further allows to fuse data based on their estimated quality. Our system comprises local and distributed heuristics to estimate the quality of sensor data, with a focus on data accuracy and data consistency. In the fusion step, the most plausible value of the measured quantity is inferred from multiple sensor readings by use of the Dempster-Shafer theory of evidence. Both quality assessment and data fusion are carried out within the network and thus do not rely on a powerful sink node. We demonstrate the effectiveness of our system by means of a wireless game controller for the game Pong, built from multiple sensor nodes. The controller can detect and reject incorrect sensor readings and thus improve the player's control over the in-game paddle.

1. Introduction

Wireless sensor networks are comprised of cheap, small, resource constrained sensor nodes that communicate using radio transceivers [1]. The sensor nodes, which typically run on batteries, are equipped with sensors to collect information about their environment. A variety of applications has emerged, as diverse as volcano monitoring, health monitoring, or fence monitoring.

Most WSN applications have in common that they depend on high quality data being reported by the sensor nodes. Incorrect, outdated, or incomplete data may have serious consequences in WSNs, like e.g. triggering an action based on incorrect data obtained from the network. However, low data quality is prevalent in WSNs. It occurs in many forms (such as spurious readings, faulty data on component failure, or lost data) and has many different causes (such as imprecision inherent to the sampling process, unstable voltage supply, or unreliable wireless communication.)

Previous approaches to the data quality problem often focus on only one facet of quality (e.g. detection of outliers), or rely on a central node for processing and cleaning [2], [3]. To solve this, we are developing an extensible system that allows to assess the quality of sensor data and allows to fuse data based on their quality. The system is designed to perform all quality assessment and data fusion tasks within the network and thus does not rely on a powerful sink node.

Our system consists of two major components: A set of heuristics and an inference engine. Both local and distributed heuristics are used to estimate the quality of either an individual sensor reading or a stream of samples. Heuristics are weighted to account for an application's individual data quality requirements. The development of application-specific heuristics is encouraged by providing an API.

The inference engine allows to fuse several correlated sensor readings by use of the Dempster-Shafer theory of evidence, a mathematical theory for reasoning in situations of uncertain knowledge [4], [5]. Since the theory does not follow a strict probabilistic interpretation [6], our system does not rely on large data sets to derive a priori distributions, like Bayesian approaches do. In contrast, the most likely value of the measured quantity is inferred from a number of correlated samples based on their estimated quality.

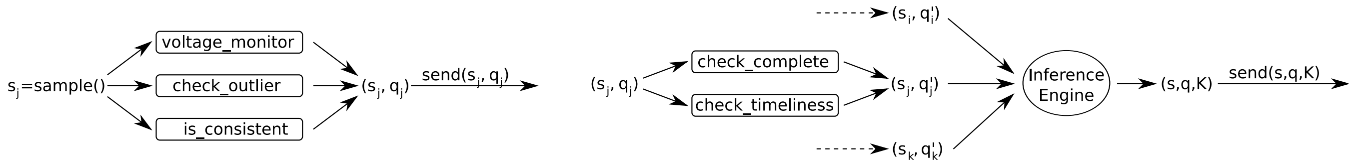
We have implemented the system on the ScatterWeb MSB430 sensor node platform [7]. Our initial experimental evaluation shows that our system effectively increases the data quality in the presence of errors. For demonstration purposes, we have built a wireless game controller for the game Pong. The use of our system improves the playability, because incorrect sensor data are detected and rejected.

2. Data Quality

The term "data quality" is used with different meanings in WSN literature. We have identified four core components of data quality, on which our system is based:

- **Accuracy:** The *accuracy* of a single sample reflects the numerical difference between the sample and the true value of the measurand. This explicitly includes errors introduced at sensor level. Intermediate processing such as aggregation may introduce further inaccuracy.
- **Consistency:** If a single sample or a stream is compliant with a user-defined model, it is *consistent*. The

978-1-4244-5113-5/09/$25.00 ©2009 IEEE



(a) Sensor Node

(b) Cluster Head

Figure 1: Data Flow (The shown heuristics are exemplary.)

model, which is dependent on the specific application, may incorporate data from multiple sensor nodes.

- **Timeliness:** The *timeliness* of data denotes if data is being received by a sink or actuator node in time. Timeliness is mainly affected by unreliable radio communication and network latency.
- **Completeness:** *Completeness* is a property of a stream; it reflects if a node has taken a sufficient number of samples to reconstruct the measurand, or if a node has successfully received a sufficient fraction of a stream from the network. Completeness is affected by sensor hardware, the chosen sampling rate, and unreliable radio communication.

Note that perfect accuracy and a correct model would always imply consistency. However, we have found consistency checks useful in the case of systematic error at sensor level, which may remain undiscovered by accuracy-related heuristics.

3. System Design

We assume that a network that uses our system is organized in clusters and sampling proceeds in rounds. One sensor node in each cluster is chosen to be the cluster head. Figure 1 shows the flow of data: In each round, every sensor node takes a sample and uses heuristics to estimate the sample's quality with respect to accuracy, consistency, and completeness. It then sends the result to its cluster head, as shown in Fig. 1a. In this example, three of the system's built-in heuristics are used, denoted by rectangles. The result is a tuple containing a sample value s and a quality estimation q. The cluster head may use heuristics to re-evaluate the quality of received data to reflect timeliness and completeness. It then fuses the samples received from the nodes in its cluster and sends the fusion result to a sink node, along with a measure of conflict K, as explained in Sec. 3.2. This process is shown in Fig. 1b.

3.1. Heuristics

A heuristic estimates the quality of a sample or a stream with respect to one or more of the above mentioned components of data quality. We use heuristics because the true value of the measurand is not known in general, and thus the accuracy of a sample can only be estimated. In our system, heuristics are functions that map an array of samples on a value of the interval $[0, 1]$, where a value of 0 denotes the lowest quality and a value of 1 denotes the highest quality. The parameter array represents the most recent sensor readings. As shown in Fig. 1, heuristics are evaluated for every sample that is taken or received from the network. The influence of each heuristic can be set by the application programmer to reflect the application's individual data quality requirements. The minimal value returned by all heuristics determines the data quality.

Our system includes a range of built-in heuristics, such as a voltage monitor, which checks the battery voltage, or a distributed outlier detection. However, the use of application-specific heuristics is encouraged by providing an API.

3.2. Inference Engine

The cluster head fuses samples received from other nodes in the cluster by use of the Dempster-Shafer theory of evidence. This is a mathematical theory that allows to infer the most plausible proposition of a set of propositions from a given set of evidences.

The theory provides a mathematical description of evidence as well as a rule for combination of evidence, called Dempster's rule of combination. A key feature is that each evidence has an associated reliability that reflects its trustworthiness. In the fusion step, the cluster head considers each received sample to be a piece of evidence supporting one of the possible true values of the measured quantity. The sample's quality is used as its reliability. By the use of Dempster's rule of combination, the most plausible value of the measured quantity is found. The application of the rule also yields the quality of the result, as well as a measure of conflict K. The latter describes the level of contradiction between the evidences, a useful information for applications that take action based on the fused data.

3.3. Initial Evaluation

For our initial evaluation, we have used four sensor nodes attached to a board to measure the board's tilt angle

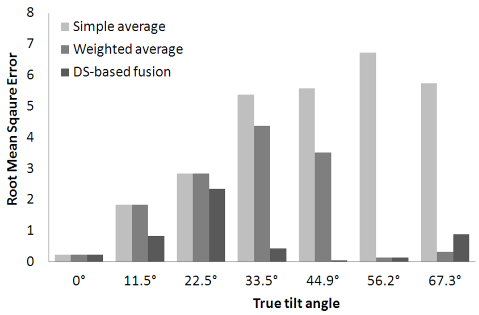
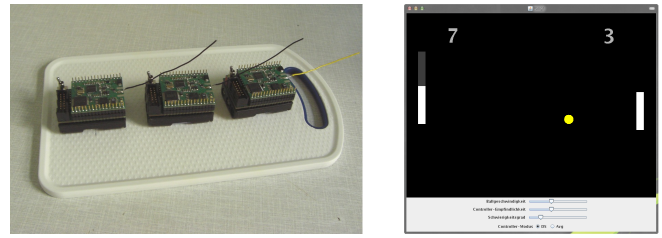


Figure 2: Root Mean Square Error of the Measured Angle



(a) The controller consists of three ScatterWeb MSB430 sensor nodes.

(b) The Game GUI

Figure 3: Wireless Game Controller for Pong

around the y-axis by using the sensors' on-board three-axis accelerometers. Two of the four nodes had been slightly twisted and thus measured an incorrect angle. Figure 2 shows the root mean square error between the true tilt angle and the angle determined with different approaches. Our fusion scheme, shown in darkest gray fill, in most cases outperforms both, the simple average approach and the weighted average approach, in which each sample is weighted by its estimated quality.

4. Demo: Wireless Game Controller

We demonstrate our system by means of a wireless game controller for the game Pong. The controller consists of three ScatterWeb MSB430 sensor nodes attached to a board, as shown in Fig. 3a. The game GUI, shown in Fig. 3b, is displayed on a laptop. By changing the controller's tilt, a player can control his in-game paddle. The other paddle is controlled by the computer.

The three sensor nodes on the controller as well as a fourth sensor node attached to the laptop form a cluster. Each controller node periodically sends tilt information that is acquired with the on-board accelerometer, along with an estimation of the quality of the tilt information. The sensor node attached to the laptop acts as the cluster head; it fuses all measurements with respect to their quality and sends the result to the laptop, where the tilt is used to determine the position of the player's paddle. The use of our system for the game controller improves playability in the presence of errors. To demonstrate this, the GUI also shows a player's "ghost paddle" (gray paddle on the left side in Fig. 3b), which is controlled using the raw sensor data. While the ghost paddle reacts to sensor errors by moving around uncontrollably, the paddle controlled with the system (shown in white on the left side in Fig. 3b) behaves as intended, because erroneous measurements are successfully detected and rejected.

The demo requires a laptop, the wireless game controller, and a cluster head node, all of which we will provide. A desk for the laptop is needed with a close-by power outlet. The sensor nodes will communicate at a frequency of 809.5 MHz, but can be reprogrammed to use another frequency if required. The setup time is about five minutes.

5. Conclusion

In this paper, we have outlined our system for quality-based data fusion in WSNs. It is centered around a set of heuristics and an inference engine that is based on the Dempster-Shafer theory of evidence. We demonstrate the effectiveness of the system by means of a wireless game controller.

Future work includes research of appropriate algorithms to build clusters and select cluster heads, as well as an in-depth evaluation.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Computer Networks*, vol. 38, no. 4, pp. 393 – 422, 2002.

[2] Zhuang, L. Chen, X. S. Wang, and J. Lian, "A weighted moving average-based approach for cleaning sensor data," in *Proceedings of the 27th International Conference on Distributed Computing Systems, 2007. ICDCS '07.*, 2007.

[3] N. Khoussainova, M. Balazinska, and D. Suciu, "Towards correcting input data errors probabilistically using integrity constraints," in *Proceedings of the 5th ACM international workshop on Data engineering for wireless and mobile access*. Chicago, Illinois, USA: ACM, 2006, pp. 43–50.

[4] A. P. Dempster, "A generalization of bayesian inference," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 30, no. 2, pp. 205–247, 1968.

[5] G. Shafer, *A mathematical theory of evidence*. Princeton University Press, 1976.

[6] D. M. Gabbay and P. Smets, Eds., *Handbook of Defeasible Reasoning and Uncertainty Management Systems, Vol. 1*, vol. 1. Kluwer, Doordrecht, The Netherlands, 1998.

[7] J. Schiller, A. Liers, H. Ritter, R. Winter, and T. Voigt, "ScatterWeb - low power sensor nodes and energy aware routing," in *Proceedings of the 38th Annual Hawaii International Conference on System Sciences, 2005. HICSS '05.*, 2005, p. 286c.